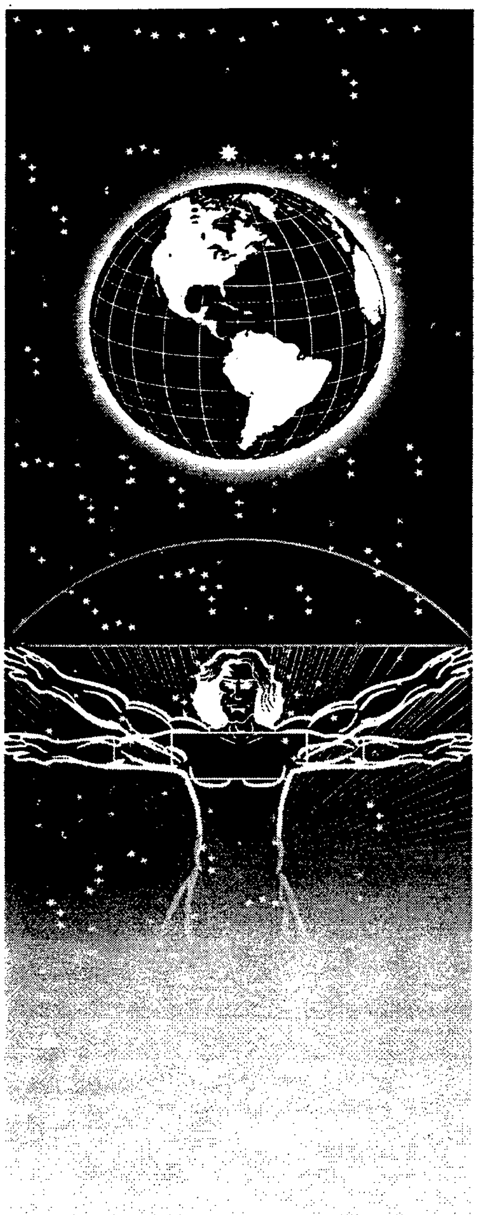
AL/EQ-TR-1997-0018

UNITED STATES AIR FORCE
ARMSTRONG LABORATORY



Review of Cost-Constrained Minimum Runs Algorithm for Response Sensitivity Analysis of the FRACT3DVS Code

Marsh Hardy

APPLIED RESEARCH ASSOCIATES, INC.
4300 San Mateo Boulevard, NE, Suite A220
Albuquerque, New Mexico 87110

May,1997

Approved for public release; distribution is unlimited.

Environics Directorate
Environmental Risk
Management Division
139 Barnes Drive, Suite 2
Tyndall Air Force Base FL
32403-5323

NOTICES

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any employees make any warranty, expressed or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness or any privately owned rights. Reference herein to any specific commercial process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency, contractor, or subcontractor thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency, contractor, or subcontractor thereof.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligations, whatsoever. The fact that the Government may have formulated or in any way supplies the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder or any person or corporation; or as conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

This technical report has been reviewed by the Public Affairs Office (PA) and is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

DENNIS O'SULLIVAN, Lt, USAF, BSC
Project Officer

Allan M Weiner

ALLAN M. WEINER, Lt Col, USAF
Chief, Environmental Risk Management Division

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1997	Final Report July 1995 to April 1997

4. TITLE AND SUBTITLE
Review of Cost-Constrained, Minimum Runs Algorithm for Response Sensitivity Analysis of the FRACT3DVS Code

5. FUNDING NUMBERS
F08635-93-C-0020

6. AUTHOR(S)
Marsh Hardy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Applied Research Associates, Inc.
4300 San Mateo Blvd., NE, Suite A220
Albuquerque, NM 87110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESSES(ES)
Armstrong Laboratory Environics Directorate
Environmental Risk Management Division
139 Barnes Drive, Suite 2
Tyndall Air Force Base, Florida 32403-5323

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AL/EQ-TR-1997-0018

11. SUPPLEMENTARY NOTES
Contracting Officer's Technical Rep: Lt Dennis O'Sullivan, (904) 283-6239

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release. Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Using the multistage design of experiments to perform a limited number of FRACT3DVS computations and stepwise regression, simplified equations are developed to predict the behavior of a funnel and gate system. The four factors included in these equation were identified to be highly significant in predicting behavior of the funnel and gate system. The equations will be useful for evaluating different funnel and gate designs and to make quick predictions in the field (a full FRACT3DVS calculation can take 1-2 days.)

The following conclusions assume the accuracy of the FRACT3DVS code and the comprehensiveness of the four parameters and two response measures studied. (1) Collectively, the first three FRACT3DVS parameters, Kaquifer/Kgate, wf/df, 2*wf/wg, are particularly important in predicting the funnel and gate's operation, especially because of the product of wf/df and 2*wf/wg. (2) It is possible to predict the operation of the funnel and gate system for ground pollution treatment with a high degree of accuracy, using a small, fast running code on a PC.

14. SUBJECT TERMS
FRACT3DVS, Funnel-and-Gate

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

(The reverse of this page is blank)

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18
298-102

PREFACE

The work described in this report covers the contract period of July 1995 through April 1996. This work was performed by Applied Research Associates (ARA), Inc., under Contract F08635-93-C-0020, Subtask 8.03, U.S. Air Force AL/EWQ, Barnes Drive, Suite 2, Tyndall Air Force Base, Florida. During the course of this study, there were two project officers, Major Mark H. Smith and Captain Jeff Stinson, BSC. This work was performed under the technical guidance of Mr. Robert E. Walker, ARA. This report was written by Marsh Hardy of ARA and edited by Mr. Walker.

EXECUTIVE SUMMARY

Funnel-and-gate systems are used for in situ groundwater remediation. To design such systems in a cost-effective manner, it is necessary to perform extensive calculations to determine the optimum system dimensions. For performing a good first-order design, a statistically sound model was developed.

Using the multistage design of experiments to perform a limited number of FRACT3DVS computations and stepwise regression, simplified equations are developed to predict the behavior of a funnel and gate system. The four factors included in these equation were identified to be highly significant in predicting behavior of the funnel and gate system. The equations will be useful for evaluating different funnel and gate designs and to make quick predictions in the field (a full FRACT3DVS calculation can take 1-2 days.)

The following conclusions assume the accuracy of the FRACT3DVS code and the comprehensiveness of the four parameters and two response measures studied. (1) Collectively, the first three FRACT3DVS parameters, Kaquifer/Kgate, wf/df, 2*wf/wg, are particularly important in predicting the funnel and gate's operation, especially because of the product of wf/df and 2*wf/wg. (2) It is possible to predict the operation of the funnel and gate system for ground pollution treatment with a high degree of accuracy, using a small, fast-running code on a PC.

(The reverse of this page is blank)

TABLE OF CONTENTS

(The reverse of this page is blank)

LIST OF FIGURES

(The reverse of this page is blank)

LIST OF TABLES

(The reverse of this page is blank)

SECTION I

INTRODUCTION

OBJECTIVE

The objective of this effort is to assess the relative contributions of nine variables that affect the FRACT3DVS simulations of a funnel-and-gate system and model those effects in a second, fast-running code.

BACKGROUND

FRACT3DVS is a large computer program used to simulate groundwater flow. A funnel-and-gate system has been devised to treat soil pollutants in situ, that is, in the field. FRACT3DVS can estimate the amount of pollution captured and flow rates in a funnel and gate system. However, the code is too computer intensive (each run can take 1-2 days on a work station) to be used for the parameter studies needed to design an optimal funnel and gate system. Hence, a simpler, faster running code is required.

SCOPE

The basic scope is to perform FRACT3DVS simulations with different values of the input variables. Using these results we can fit a simple function (i. e., a response surface) that attempts to match the FRACT3DVS output for each set of inputs. Each run of FRACT3DVS takes 1-2 days on a minicomputer or work station and has to be performed off-site by an expert in the use of the code. To get the most information in the least number of runs, it was decided to use a multistage design of experiments. This multistage approach allows us to reduce the number of variables and, therefore, reduce the required number of FRACT3DVS runs. The multistage approach involves a set of "screening runs," followed by additional runs to better define the response sensitivities. "Analysis of variance" (1-3,12) (ANOVA), a statistical method to measure the relative importance of terms in linear models, was used first to find significant linear effects in the screening runs, then later to find overall models of the funnel-and-gate system response measures. Two of the code's output measures were chosen to produce response surface models of the code: relative area and relative flow, in that order of importance.

SECTION II

STAGE 1 -- SCREENING

A complete, quadratic response model in nine variables has 55 terms and would need at least that many runs in order to estimate all of the parameters. Since any significant high order terms are more likely to involve variables that have significant linear effects than those that do not, it is useful to try to reduce the number of variables via screening runs and an ANOVA of the data generated by the screening runs.

A Plackett-Burman (4,5,10,12) experimental design was used for the screening stage of initial runs of the code. Plackett-Burman designs are 2-level designs, run at typical high and low values for each variable. For k variables, Plackett-Burman designs require 4*(int(k/4)+1) runs, i.e., the least multiple of 4, greater than k, which in this case meant 12 runs. The subsequent ANOVA needs at least an additional run, with all variables held at their nominal or central values, to create some model error to use for comparison purposes in the significance tests of the linear response terms. Hence, 13 runs of FRACT3DVS are necessary for the nine factors considered for that screening analysis.

The screening design above was computed for the purpose of identifying important linear effects of the nine FRACT3DVS parameters in the code's simulation of a funnel and gate system. An advantage of the Plackett-Burman design used for the initial screening phase of this design is that it requires fewer than the usual 2*k+1 runs needed for a “one at a time plus center point” design, which would have been 19 runs for this case.

Accordingly, the first experimental design resulted in the X matrix shown in Table 1. The X matrix defines the series of runs given by the Plackett-Burman design with center point, for an initial screening of the nine FRACT3DVS parameters. The results of that screening are used later to compute the next series of runs to perform. Note, the matrix is coded so that -1=minimal, 0=medial, and 1=maximal values.

TABLE 1. FIRST SCREENING DESIGN.

Run	Kh	Hydraulic Gradient	Funnel Width	Funnel Depth	Funnel Thickness	Funnel Kh	Gate Height	Gate Width	Dynamic Reaction
1	1	-1	1	-1	-1	-1	1	1	1
2	1	1	-1	1	-1	-1	-1	1	1
3	-1	1	1	-1	1	-1	-1	-1	1
4	1	-1	1	1	-1	1	-1	-1	-1
5	1	1	-1	1	1	-1	1	-1	-1
6	1	1	1	-1	1	1	-1	1	-1
7	-1	1	1	1	-1	1	1	-1	1
8	-1	-1	1	1	1	-1	1	1	-1
9	-1	-1	-1	1	1	1	-1	1	1
10	1	-1	-1	-1	1	1	1	-1	1
11	-1	1	-1	-1	-1	1	1	1	-1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	0	0	0	0	0	0	0	0	0

As it turned out, the matrix X given in Table 1 was not entirely representative of the funnel and gate process to be simulated by the FRACT3DVS code. It was not physically consistent to combine high and low values for all pairs of parameters, which is to say that some pairs were correlated.

After a review of this matrix and of the process to be simulated, the nine original parameters were reduced to four independent parameters. The ratio of Kh to funnel Kh became "Kaquifer/Kgate". The ratio of funnel width to funnel depth became "wf/df." The ratio of 2*funnel width to gate width became "2*wf/wg," and hydraulic gradient was now to be called "del_h/del_l." Funnel thickness was deemed to be minimally important to irrelevant. Gate height was to be the same as funnel depth, and dynamic reaction, a chemical reaction, was not to be considered.

A new Plackett-Burman design matrix X, shown in Table 2, was then computed. Note, the nine runs here happen to be the same as 2*k+1, the number of runs required in a one-at-a-time analysis. Plackett-Burman's advantage of fewer number of runs occurs when there are more than 4 variables. Nevertheless, P-B designs have another advantage over one-at-a-time; variables vary throughout the entire set of runs, each variable simultaneously in the presence of each other's variation, which can produce better models.

After the nine new screening runs had been completed the results were analyzed by ANOVA. Relative Area could be represented with a high degree of accuracy as a simple linear model in the first three of the four new parameters -- so accurate that the model explained 97.66 percent of the variation of Relative Area in the screening stage (and which can be expected to change with new, augmenting data).

TABLE 2. PLACKETT-BURMAN INITIAL SCREENING DESIGN WITH CENTER POINT FOR SENSITIVITY ANALYSIS OF 4 FRACT3DVS PARAMETERS.

Run	Kaquifer/Kgate	wf/df	2*wf/wg	del_h/del_l	Note:
1	4	2.0	10	0.005	(Case 1)
2	1	0.5	5	0.001	P-B
3	1	0.5	20	0.010	P-B
4	1	4.0	5	0.010	P-B
5	1	4.0	20	0.001	P-B
6	20	0.5	5	0.010	P-B
7	20	0.5	20	0.001	P-B
8	20	4.0	5	0.001	P-B
9	20	4.0	20	0.010	P-B

SECTION III

STAGES 2, 3, AND 4

Augmentation of the screening design was reported in "stages" of new runs of decreasing necessity or priority, as shown in Table 3. Stage 1, the screening design was "D-optimally" augmented into a second, higher resolution design, able to estimate second order terms of the significant variables identified by the ANOVA. The D-optimality criterion (5-8) seeks to maximize $|X'X|$, the determinant of the design's information matrix. For this operation, the non-significant variables are held at their nominal values.

In general, the second, augmenting stage can be divided into second and third stages that first use a 2-level then a 3-level design with ANOVAs after each before producing the final response surface. Here, the low number of variables allowed calculating all augmenting runs after one ANOVA. Stage 2 allowed estimation of all 2nd order terms of the 3 significant variables evident in Stage 1. Stage 3 allowed estimation of the remaining 2nd order terms involving the 4th, so far non-significant variable, "grad h" (del_h/del_l). Stage 4, corresponds to a term for the product of all 4 parameters and ensures a degree of an error lack-of-fit and avoids the problem of an “over-determined” model.

Normalizing transformations of the data can be employed to improve model fit, or transformations can be used to guarantee appropriate bounds on the response model's estimated values. Here, log transforms kept the output values positive.

The seven augmenting runs (10-16) are shown in the augmented X matrix above. A full, quadratic, response surface model was fit to the 16 FRACT3DVS runs, but the analysis based on that model was unreliable since that model's parameter estimates were not, themselves, reliable. A reduced, quadratic, response surface model was required, and one was found using stepwise regression. Its ANOVA table follows showing the model form and its improved parameter estimates.

The reduced model has higher-order terms in second and third parameters only, wf/df and 2*wf/wg, so only those two needed to be coded for a response surface analysis. (The two coding formulae are shown in Table 4.) An analysis of the surface shows a conditional

maximum at wf/df = .3191066 and 2*wf/wg = 4.2285113, shown in Figure 1. The response surface analysis tells us we can maximize the Relative Area measure of performance by designing the funnel and gate system to have parameters as close to these values as is practical to do so. "Canonical" and "ridge" analyses (3,9-12) were two procedures used to find this singularity point.

TABLE 3. AUGMENTED, D-OPTIMAL, SATURATED, 3-LEVEL DESIGN + ERROR.

Stage	Run	Kqauifer/Kgate	wf/df	2*wf/wg	del_h/del_l	Relative Area
1	1	4	2.0	10	0.005	0.280
	2	1	0.5	5	0.001	0.379
	3	1	0.5	20	0.010	0.320
	4	1	4.0	5	0.010	0.250
	5	1	4.0	20	0.001	0.155
	6	20	0.5	5	0.010	0.457
	7	20	0.5	20	0.001	0.353
	8	20	4.0	5	0.001	0.285
	9	20	4.0	20	0.010	0.182
2	10	4	4.0	20	0.005	.
	11	20	4.0	10	0.005	.
3	12	1	4.0	5	0.001	.
	13	4	0.5	10	0.010	.
	14	4	0.5	20	0.001	.
	15	20	0.5	10	0.001	.
4	16	1	0.5	5	0.010	.

The first and fourth FRACT3DVS parameters, Kaquifer/Kgate and del_h/del_l, have monotonically increasing, positively correlated effects on relative area, so no global maximum exists. The higher they go, the greater relative area will be.

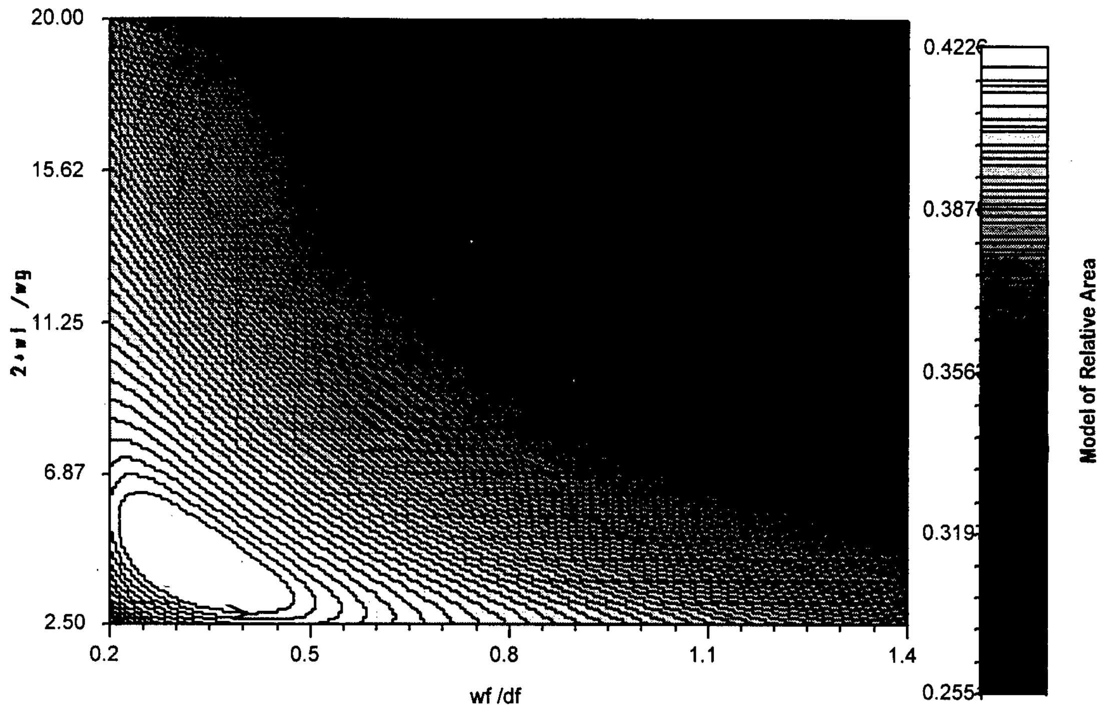


Figure 1. Quadratic Response Surface for Relative Area in Two FRACT3DVS Parameters. (The two other parameters are held at the means of their natural logs.)

The multiple linear regression model for the natural log of relative area and its estimated coefficients are shown in Table 4, the ANOVA table for the model. Table 4 also contains the T-statistics associated with each model parameter. The square of the T statistic is a good measure of the relative importance of each term in the model. This measure is proportional to the variance of the variable in the term and the square of the term's estimated coefficient. The only terms allowed in the model were those that passed the T test at a 0.05 level.

TABLE 4. LINEAR MODEL'S DEPENDENT VARIABLE: LN (RELATIVE AREA).

Analysis of Variance (measures and tests for goodness of fit)

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	7	1.66818	0.23831	736.242	0.0001
Error	8	0.00259	0.00032		
C Total	15	1.67077			

Root MSE	0.01799	R-square	0.9985
Dep Mean	-1.25228	Adj R-sq	0.9971
C.V.	-1.43669		

Parameter Estimates (with significance tests)

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
Intercept	1	-1.155717	0.03121130	-37.029	0.0001
ln1	1	0.049583	0.00379650	13.060	0.0001
cd2	1	-0.288244	0.00470598	-61.251	0.0001
cd3	1	-0.167119	0.00527006	-31.711	0.0001
ln4	1	0.015426	0.00448120	3.442	0.0088
cd2cd3	1	-0.062542	0.00539519	-11.592	0.0001
cd2cd2	1	-0.073530	0.02414147	-3.046	0.0159
cd3cd3	1	-0.031231	0.01271146	-2.457	0.0395

Name	Definition
ln1	ln(Kaquifer/Kgate)
cd2	coded ln(wf/df) = (ln(wf/df) -0.346574) / 1.039721
cd3	coded ln(2*wf/wg) = (ln(2*wf/wg) -2.302585) / 0.693147
ln4	ln(del_h/del_l)
cd2cd3	cd2 * cd3
cd2cd2	cd2 * cd2
cd3cd3	cd3 * cd3

Note: A 95% confidence bound for a value of Relative Area predicted by this eq. is (predicted value *or/ 1.0423575), where 1.0423575=exp(t(8,.975)*Root MSE).

SECTION IV

RELATIVE FLOW

Since all analysis and modeling decisions had been based on ANOVAs of relative area, the concern was expressed as to how to transform the results to a response model of relative Qgate (a measure of relative flow). Relative Qgate tended to vary with relative area but not exactly, as shown in Figure 2, which displays a cubic trend with 95 percent bounds for the trend line.

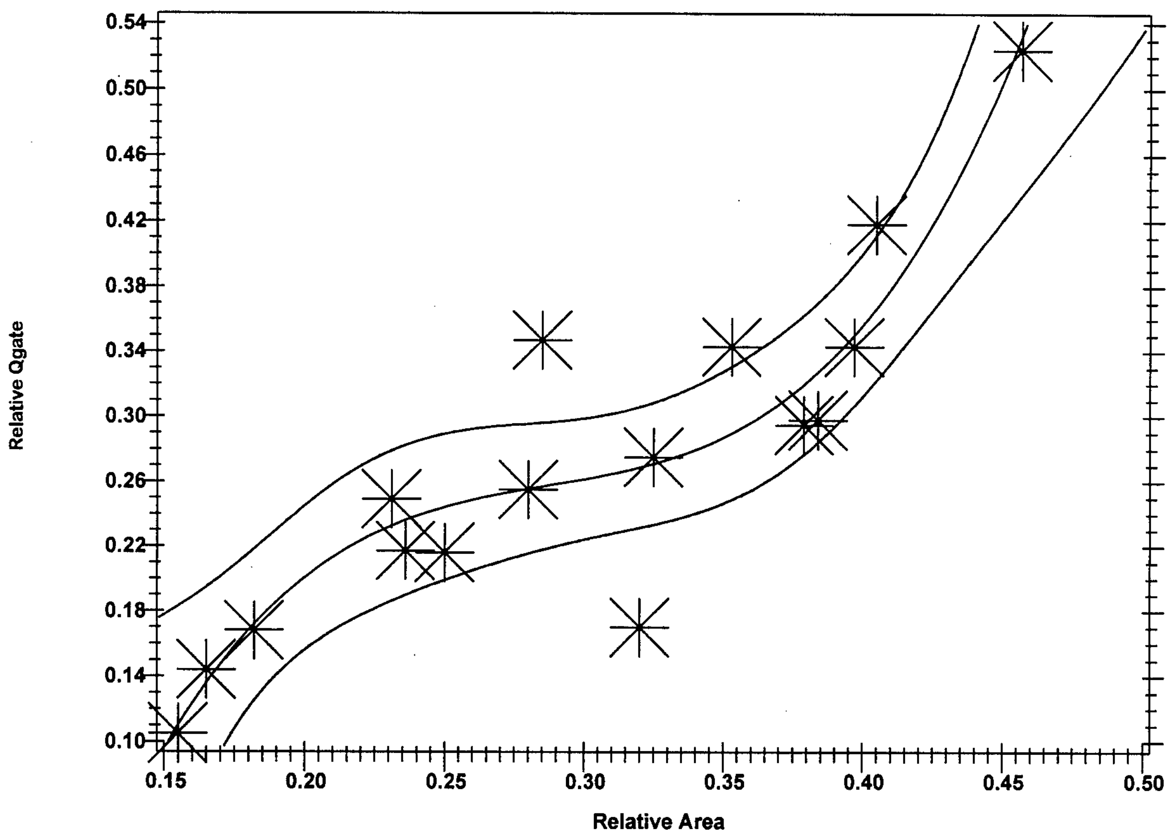


Figure 2. Relative Qgate Trend in Relative Area.

As with relative area, a full, quadratic, response surface model was fit to the 16 FRACT3DVS runs, but an analysis based on that model was unreliable since that model's

parameter estimates were not, themselves, reliable. A reduced, quadratic, response surface model was required, and one was found using stepwise regression. The model for relative Qgate fit the data at least as well as the model for relative area. This was fortuitous, since the decisions as to which variables to concentrate on were based on modeling relative area. The flow model ANOVA, Table 5, follows showing model form and coefficient estimates. As in Table 4, model terms were selected at a 95 percent confidence level.

The reduced model has higher order terms in all 4 FRACT3DVS parameters and linear terms in the 1st and 3rd parameters only, Kaquifer/Kgate and 2*wf/wg. Such a model cannot be formally analyzed for regions of optimality by the SAS response surface procedure available and was not attempted herein. Running this model with a fine 4D mesh is another way to find regions of optimality, and the results can be presented graphically. This could be a topic of some future investigation, but we know relative area and relative flow increase together.

TABLE 5. LINEAR MODEL'S DEPENDENT VARIABLE: LN(RELATIVE QGATE).

Analysis of Variance (measures and tests for goodness of fit)

Source	df	Sum of Squares	Mean Square	F Value	Prob>F
Model	8	2.58870	0.32359	1494.499	0.0001
Error	7	0.00152	0.00022		
C Total	15	2.59022			

Root MSE	0.01471	R-square	0.9994
Dep Mean	-1.37525	Adj R-sq	0.9987
C.V.	-1.06995		

Parameter Estimates (with significance tests)

Variable	Definition	df	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
intercept		1	-0.484772	0.05102584	-9.501	0.0001
ln1	ln(Kaquifer/Kgate)	1	0.332230	0.01618190	20.531	0.0001
ln2	ln(wf/df)	.	(0.)			
ln3	ln(2*wf/wg)	1	-0.575412	0.04126705	-13.944	0.0001
ln4	ln(del_h/del_l)	.	(0.)			
ln1ln2	ln1 * ln2	1	-0.013979	0.00339552	-4.117	0.0045
ln1ln3	ln1 * ln3	1	0.015540	0.00468619	3.316	0.0128
ln2ln3	ln2 * ln3	1	-0.090229	0.00243975	-36.983	0.0001
ln3ln4	ln3 * ln4	1	-0.024498	0.00697718	-3.511	0.0098
ln1ln1	ln1 * ln1	1	-0.059383	0.00415047	-14.308	0.0001
ln4ln4	ln4 * ln4	1	-0.003971	0.00135214	-2.937	0.0218

Note: A 95% confidence bound for a value of Relative Flow predicted by this eq. is (predicted value *or/ 1.0354014), where 1.0354014=exp(t(7,.975)*Root MSE).

SECTION V

SUMMARY AND CONCLUSIONS

Using the multistage design of experiments to perform a limited number of FRACT3DVS computations and stepwise regression, simplified equations are developed to predict the behavior of a funnel and gate system. The four factors included in these equation were identified to be highly significant in predicting behavior of the funnel and gate system. The equations will be useful for evaluating different funnel and gate designs and to make quick predictions in the field (a full FRACT3DVS calculation can take 1-2 days.)

The following conclusions assume the accuracy of the FRACT3DVS code and the comprehensiveness of the four parameters and two response measures studied. (1) Collectively, the first three FRACT3DVS parameters, Kaquifer/Kgate, wf/df, 2*wf/wg, are particularly important in predicting the funnel and gate's operation, especially because of the product of wf/df and 2*wf/wg. (2) It is possible to predict the operation of the funnel and gate system for ground pollution treatment with a high degree of accuracy, using a small, fast-running code on a PC.

REFERENCES

1. Hocking, R. R., "The Analysis and Selection of Variables in Linear Regression," Biometrics, 32, 1976.

2. Draper, N. and H. Smith, Applied Regression Analysis, 2d Edition, New York: John Wiley & Sons, Inc., 1981.

3. SAS Institute Inc., SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989.

4. Plackett, R. L. and J. P. Burman, "The Design of Optimum Multifactorial Experiments," Biometrika, *33*, 1947.

5. SAS Institute Inc., SAS/QC Software: Usage and Reference, Version 6, First Edition, Volumes 1 and 2, Cary, NC: SAS Institute Inc., 1995.

6. Dykstra, O., Jr., "The Augmentation of Experimental Data to Maximize |X'X|," Technometrics, 13, 1971.

7. Fedorov, V. V., Theory of Optimal Experiments, translated and edited by W. J. Studden and E. M. Klimko, New York: Academic Press, 1972.

8. Cook, R. D. and C. J. Nachtsheim, "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," Technometrics, 22, 1980.

9. Draper, N. R., "Ridge Analysis of Response Surfaces," Technometrics, 5, 1963.

10. John, P. W. M., Statistical Design and Analysis of Experiments, New York:Macmillan Publishing Co., Inc., 1971.

11. Myers, R. H., Response Surface Methodology, Blacksburg, VA: Virginia Polytechnic Institute and State University, 1976.

12. Box, G. E. P. and J. S. Hunter, Statistics for Experimenters, New York: John Wiley & Sons, Inc., 1978.